

## DISEÑO E IMPLEMENTACIÓN DE UN CAPTURADOR DE DATOS WEB

Plazas G. David, Romero R. Henry

**Resumen:** El objetivo del presente artículo es el diseño de un capturador de datos web que a través de un circuito electrónico se conecta a la internet, funcionando independientemente de un computador para acceder a un dato variable desde una página web. Aprovechando la tecnología del hardware libre a través del módulo Arduino y la del software libre mediante el cual se programa y se comprueba la aplicación. Se realizaron pruebas donde se mostraron las funcionalidades de cada dispositivo utilizado, y de la integración del capturador de datos web bajo diferentes ambientes de trabajo y conexión. El producto final es un equipo de uso académico para demostrar el uso de herramientas del Software Libre y dispositivos de Hardware Libre en aplicaciones orientadas a protocolos TCP/IP, HTTP, UDP y la electrónica en general.

**Palabras Clave:** Software Libre/ Hardware Libre/ Arduino/ WEB/ TCP/IP

### DESIGN AND IMPLEMENTATION OF A WEB DATA CAPTURER.

**Abstract:** The aim of this paper is to design a web data capturer through an electronic circuit connected to the INTERNET using hardware and software standards. The main goal is to make it work independently from a computer to access a variable data from a website on the Internet. This device relies on free hardware technology through the Arduino module and free software used to program the device and tryouts. The tests conducted showed the features of each device used and the integration of the web data capturer under different work environments and connections. The final product is a device developed for academic purposes in order to show free software and free hardware technology oriented to TCP/IP, HTTP, and UDP protocols and general electronics as well.

**Keywords:** Free Software/ Free Hardware/ Arduino/ WEB/ TCP / IP

## I.- INTRODUCCIÓN

Si el usuario necesita obtener un simple dato publicado en una página web específica, como la temperatura de la ciudad, los puntos que subieron o bajaron sus acciones en la bolsa de valores, en que lugar va su equipo de béisbol en la temporada, el nivel de un silo en su empresa, la cantidad de productos elaborados ó vendidos en un periodo de tiempo determinado, no es necesario que tenga al alcance un ordenador personal, ya que, el mismo resulta un objeto muy costoso, que ocupa espacio, consume mucha energía y tarda demasiado para acceder a un dato a través del internet (entre el encendido, ejecución de la herramienta de búsqueda o el navegador web y el apagado), por lo que se debe conseguir una salida más productiva con un dispositivo electrónico económico de poco tamaño, que permita ser configurado para acceder rápidamente a internet con el único propósito de buscar la información que el usuario necesite en tiempo real.

Es por esta razón que el objetivo general es el de diseñar e implementar un capturador de datos web que no requiere de un computador para acceder a éstos, basándose en tecnología bajo licenciamiento del Software Libre, módulos de captura, conversión y procesamiento de datos.

## II.- DESARROLLO

En el desarrollo de la investigación se definieron los métodos que guiarían el diseño del dispositivo, luego se identificaron los componentes necesarios para la creación del prototipo para luego evaluar los resultados del mismo.

### 1 El Diseño

Para poder realizar el capturador de datos web, se comenzó por diseñar el dispositivo electrónico capaz de enviar y recibir información a través de una red Ethernet de manera autónoma, para esto, se cumplieron con las fases descritas a continuación.

### 2 Fases del Diseño

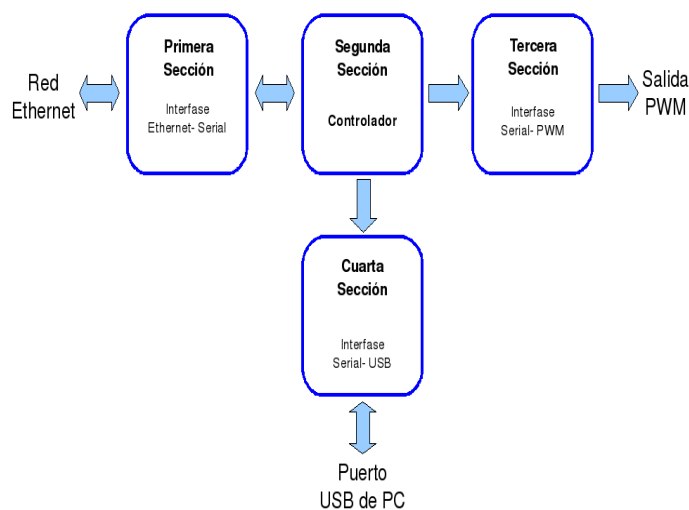
En el desarrollo del diseño del capturador de datos web se decidió dividir el trabajo en varias fases, con la intención de facilitar el cumplimiento de los objetivos planteados, a continuación se listan las fases que conforman el diseño:

- Realización del diseño conceptual del hardware.
- Realización del diagrama de bloques de acuerdo al modelo TCP/IP.
- Funcionamiento general del circuito.
- Resultados.

## 2.1

### Realización del Diseño Conceptual del Hardware.

Desde el punto de vista conceptual, el diseño del capturador de datos web consta de cuatro (04) secciones principales que dependen entre si para poder acceder a la red, buscar la página web con el dato deseado de la temperatura, tomar el dato, transformar ese dato con la técnica de PWM para que el usuario final pueda observar el valor de la temperatura y llevar un seguimiento de como se va ejecutando cada uno de estos pasos. ver Figura 1.



**Figura 1 Diagrama Conceptual del Hardware del Capturador de Datos Web**

En la primera sección ocurre una conversión entre el protocolo Ethernet (que permite el acceso a internet) y el protocolo de comunicación que utilice la segunda sección. Esta conversión de protocolos o interfaz ocurren de manera bidireccional puesto que habrá información y comandos viajando en ambas direcciones.

La segunda sección almacena y ejecuta las instrucciones y variables necesarias para enviar y recibir datos a la primera sección, así como transmitir el dato correcto para ser mostrado por la tercera sección.

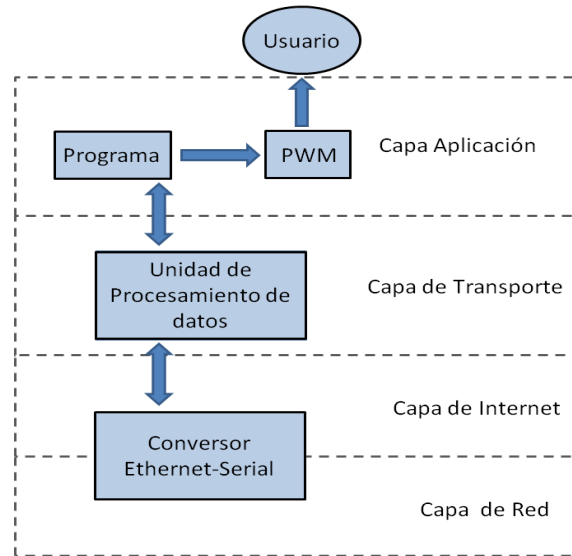
La tercera sección toma el dato en el protocolo de comunicación de la segunda sección y lo convierte a un valor de voltaje equivalente usando PWM.

La cuarta sección va directamente conectada a la segunda sección para poder capturar los datos que permitan determinar como se van desarrollando las fases de todos lo procesos que se desarrollan en el diseño.

## 2.2 Realización del Diagrama de Bloques de acuerdo al Modelo TCP/IP.

Con la intención de realizar un diseño ordenado, se ha decidido separar en diferentes fases, las cuales están basadas en el modelo estándar internacional de capas TCP/IP. Ver Figura 2.

Dicho modelo está conformado por cuatro capas, que van desde la conexión física del cableado estructurado tipo UTP, los protocolos de comunicación en lenguaje de máquina, hasta la aplicación que se comunica en el lenguaje del usuario.



**Figura 2 Diagrama de Bloques y Modelo TCP/IP Conforme al RFC1122 Y RFC1123 de la IETF.**  
**Para cumplir con las especificaciones del modelo de capas TCP/IP el diseño del dispositivo Capturador de Datos Web (CDW) se describe de la siguiente manera:**

#### **2.2.1 Capa de Red.**

El requisito de disponer de una conexión física estándar en el CDW queda satisfecho con un puerto RJ45 que es el conector más utilizado para cable UTP estructurado.

Además, Como las prácticas y pruebas del dispositivo aprovechan redes de acceso local (LAN) el protocolo de transmisión de datos es Ethernet.

#### **2.2.2 Capa de Internet.**

Siguiendo las reglas de resolución de direcciones de la segunda capa se utilizan direcciones y máscaras IP para identificar la red, el host, y la IP de difusión con el fin de determinar la dirección de origen/destino de la información. Así como la vía de salida/llegada de las solicitudes/respuestas desde y hacia la internet.

#### **2.2.3 Capa de Transporte.**

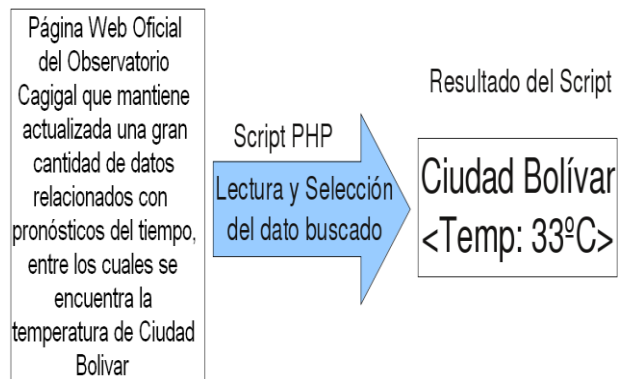
El dispositivo utiliza TCP y UDP como protocolos de transporte para controlar el flujo de datos. Esto es el estándar para las comunicaciones HTTP para clientes web establecido por la IANA (Internet Assigned Numbers Authority ó la autoridad para la asignación de números en la internet). Teniendo como puerto asignado el número 80.

#### **2.2.4 Capa de Aplicación.**

Esta capa abarca la programación del Arduino y la conversión del dato de la temperatura a PWM. Este último proceso presenta el valor de la temperatura de manera de que con la ayuda de un multímetro pueda ser entendida por el usuario final, completando así las especificaciones de las capas del modelo TCP/IP.

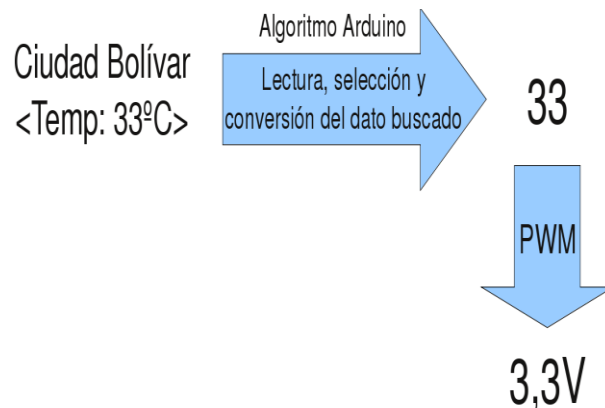
### **2.3 Funcionamiento General.**

En la Figura 3 se observa que el dato variable que va a ser capturado por el dispositivo es el de la temperatura de Ciudad Bolívar, este dato es obtenido por un script escrito en lenguaje PHP desde la página oficial del Observatorio Cagigal.



**Figura 3 Diagrama del funcionamiento del script PHP.**

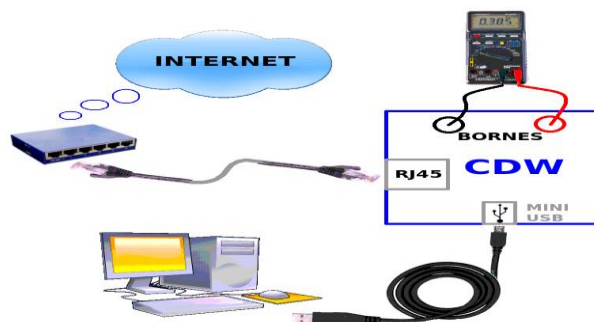
El circuito capturador de datos web, se conecta a internet y lee el script con la temperatura de Ciudad Bolívar que esta alojado en un servidor web. Luego utilizando la técnica de PWM, convierte el valor de temperatura, en un valor de voltaje proporcional. Ver Figura 4.



**Figura 4 Diagrama del funcionamiento del Circuito Electrónico.**

### 2.3.1 Conexiones y puesta en marcha:

El capturador de datos web (CDW) se conecta directamente a un módem de internet ADSL o a un router o switch con acceso a internet, paralelamente se puede (aunque no es indispensable para el funcionamiento del dispositivo) conectar el puerto USB a una computadora para observar el trabajo paso a paso de módulo arduino como dispositivo de control del CDW. Ver Figura 5.



**Figura 5 Esquema de conexiones del Capturador de Datos Web.**

Para observar el valor de temperatura capturado basta con conectar un voltímetro a los bornes del CDW para observar el valor de temperatura proporcional en voltaje (con una relación de 1 a 1/10, por ejemplo, 20°C corresponde a 2,0V).

### **2.3.2 Pruebas realizadas:**

Con el fin de verificar el funcionamiento de los componentes del circuito se realizó una prueba al módulo Arduino, y una prueba en conjunto con el módulo FT232R. Por otro lado, se realizó una prueba con el módulo Lantronix con el módulo FTDI FT232R.

Para realizar la cuarta, quinta, sexta y séptima prueba se utilizaron dos tipos de modems. El primero de ellos es fabricado por la empresa ZTE que trabaja bajo el estándar ADSL de conversión de cableado telefónico a Ethernet del proveedor CANTV.

El segundo módem es distribuido bajo la marca Axxestel y permite el acceso a internet a través de una conexión inalámbrica con el proveedor Movistar usando el protocolo dial-up.

Para garantizar que el CDW puede trabajar perfectamente en un segmento de red privado o local, se probó el funcionamiento del dispositivo al compartir una conexión a internet con dos PC de escritorio. En esta red local se utilizó un *switch* marca Encore, un *router* marca Cnet y un servidor Firewall con DHCP.

## **3 Resultados obtenidos:**

El CDW accedió al internet sin dificultad, ubicó el script desarrollado en PHP y obtuvo el valor de la temperatura de Ciudad Guayana tal como fue configurado.

El programa se ejecutó de manera satisfactoria en un periodo de 1800 segundos de prueba.

### **3.1 Discusión de Resultados:**

El capturador de datos web logra conectarse a la red de redes de forma satisfactoria puesto que cumple con los estándares internacionales para la comunicación Ethernet, de igual manera, se conecta a la página web con el valor de temperatura y realiza la lectura de la misma.

La salida PWM se configura desde la programación del módulo arduino y el valor de dos dígitos de temperatura ambiental de la ciudad se presenta correctamente mientras se disponga de una alimentación ideal.

## **III. CONCLUSIONES.**

Del presente trabajo de investigación se obtuvieron las siguientes conclusiones:

1. Se ha logrado diseñar un equipo electrónico capaz de capturar datos web desde cualquier red o subred. Funcionando correctamente con dispositivos que cumplen con los estándares internacionales para la comunicación Ethernet base 100 y base 10.
2. El equipo funciona de manera completamente automática, por lo tanto el usuario solo debe observar su funcionamiento y resultado, tal como se muestra en el manual del usuario. El juego de leds informa sobre el cumplimiento de los pasos necesarios para que el CDW cumpla con su programación, y a su vez puede servir para detectar en que paso se esté produciendo un error.
3. De forma similar la salida de debug USB permite ver por faces el cumplimiento de los protocolos de comunicación implicados, y así detectar con facilidad posibles errores tanto en el capturador de datos, como en los dispositivos que conforman la red.
4. El producto final, es un equipo de uso académico para demostrar el uso de herramientas del Software Libre y

dispositivos de Hardware Libre en aplicaciones orientadas a protocolos TCP/IP, HTTP, TCP, UDP y la electrónica en general.

#### **IV. BIBLIOGRAFÍA**

1. Rojas de N, R (1997) Orientaciones Practicas para Elaboración de Informes de Investigación (2<sup>da</sup> Ed.) UNEXPO. Vice-Rectorado Puerto Ordaz.
2. Quintana, D (2007) Diseño e Implementación de una Red de Telefonía IP con Software Libre en el Raap (Trabajo de Grado) Pontificia Universidad Católica del Perú.
3. Herlitz, H (2005) Transversabilidad en NAT/FIREWALL (Trabajo de Grado) Universidad Católica de Temuco.
4. Solano, H (2005) Interfaz para el Monitoreo de Redes de Comunicaciones mediante una Aplicación Web (Trabajo de Grado) Universidad de las Américas Puebla